

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

Requested Patent: JP6187374A
Title: RETRIEVAL SYSTEM
Abstracted Patent: JP6187374
Publication Date: 1994-07-08
Inventor(s): TAKASHIMADA MASAYA
Applicant(s): HOKKAIDO NIPPON DENKI SOFTWARE KK
Application Number: JP19920333950 19921215

Priority Number(s):
IPC Classification: G06F15/40

Equivalents:

ABSTRACT:

PURPOSE: To retrieve the target text with not only a predetermined key word, but also even a similar key word.

CONSTITUTION: At the time of registration, a text input means 111 receives the text and a text storage number acquiring means 112 stores it in a text file. A key word and attribute input means 113 receives a key word for retrieval and its attribute and stores the relation between the text storage number, and key word and its attribute in an index file 13. At the time of the retrieval, once a key word input means 121 receives the key word, a key word retrieval means 122 and an all-corresponding-text-storage-number acquiring means 123 search for the attribute on the basis of the key word and for similar key words on the basis of the attribute, the corresponding text storage number is determined through a bit map generating means 124, and the text is obtained by a corresponding text acquiring means 15.

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平6-187374

(43)公開日 平成6年(1994)7月8日

(51)Int.Cl.⁵
G 0 6 F 15/40

識別記号 庁内整理番号
5 0 0 T 7218-5L
Q 7218-5L

F I

技術表示箇所

審査請求 未請求 請求項の数1(全12頁)

(21)出願番号

特願平4-333950

(22)出願日

平成4年(1992)12月15日

(71)出願人 000241979

北海道日本電気ソフトウェア株式会社
北海道札幌市中央区南一条西4丁目5番地
1号

(72)発明者 ▲高▼島田 正哉

北海道札幌市中央区南一条西4丁目5番地
1号北海道日本電気ソフトウェア株式会社
内

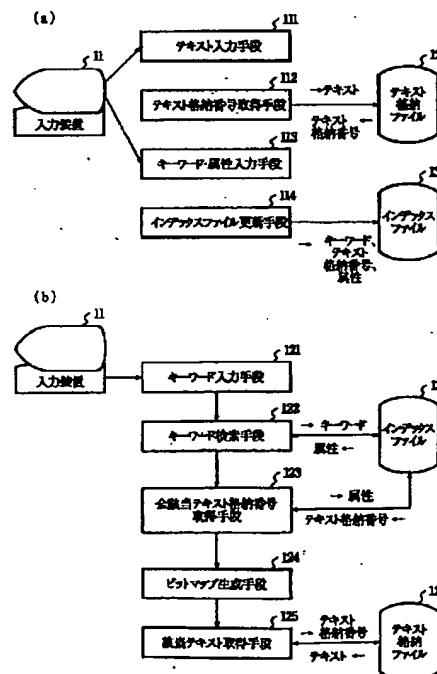
(74)代理人 弁理士 京本 直樹 (外2名)

(54)【発明の名称】 検索方式

(57)【要約】

【目的】 予め決められたキーワードだけでなく、類似したキーワードからも目的とするテキストを検索できるようとする。

【構成】 登録時には、テキスト入力手段111でテキストを受取り、テキスト格納番号取得手段112でテキストファイルに格納する。キーワード・属性入力手段113で検索のキーワードとその属性とを受取り、テキスト格納番号とキーワードと属性との関連をインデックスファイル13に格納する。検索時には、キーワード入力手段121でキーワードを受取ると、キーワード検索手段122、および全該当テキスト格納番号取得手段123で、キーワードから属性、属性から類似キーワードを検索し、該当するテキスト格納番号をピットマップ生成手段124を介して決定し、該当テキスト取得手段125によりテキストを得る。



【特許請求の範囲】

【請求項1】 キーワードを与えることにより文字列で構成されるテキストを検索する検索方式において、入力装置から文字列で構成されるテキストを受取るテキスト入力手段と、受取ったテキストをテキスト格納ファイルに格納し、そのテキストを一意に決定する番号を取得するテキスト格納番号取得手段と、入力装置から前記テキストを構成するキーワードとその属性とを受取るキーワード・属性入力手段と、このキーワード・属性入力手段が受取ったキーワードと同じ属性を有するレコードをインデックスファイルから探し出し、そのレコードからリンクされているキーワードのレコード群の最後に新たにレコードを追加し、インデックスファイルを更新するインデックスファイル更新手段と、入力装置から検索の対象となるキーワードを受取るキーワード入力手段と、このキーワード入力手段が受取ったキーワードを構成要素とするレコードを前記インデックスファイルから探し出し、そのレコードからリンクされている属性を構成要素とするレコードを取り出し、属性を取得するキーワード検索手段と、このキーワード検索手段で得られた属性をもとに、再度前記インデックスファイルを検索し、同一属性をもつレコードからリンクされている全てのキーワードレコードを探し出し、その各キーワードレコードから更にリンクされている該当テキストの格納番号を構成要素とするレコードから、テキストの格納番号の全てを取得する全該当テキスト番号取得手段と、この全該当テキスト番号取得手段が取得した全ての格納番号をビット列により表現し、論理和をとることにより該当テキストの格納番号を得るビットマップ生成手段と、このビットマップ生成手段で得られたテキスト格納番号をもとに前記テキスト格納ファイルを検索してテキストを取得する該当テキスト取得手段とを有することを特徴とする検索方式。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は検索方式、特に検索しようとするテキストのキーワードが不明であっても類似したキーワードを与えることにより目的とするテキストを検索する検索方式に関する。

【0002】

【従来の技術】 従来の検索方式は図2(a)にテキスト登録時の構成図を示すように、テキストとそれを検索するためのキーワードとを入力装置21から受取るテキスト・キーワード入力手段211と、テキストをテキスト格納ファイル22に格納し、そのテキストを一意に決定する番号を取得するテキスト格納番号取得手段212と、インデックスファイル23に存在する同一のキーワードを有するインデックスレコードに、テキスト格納番号取得手段212で得られた番号を追加し、インデックスファイル23を更新するインデックスファイル更新手段

段213とを有し、図2(b)にテキスト検索時の構成図を示すように、検索対象となるキーワードを入力装置21から受取るキーワード入力手段221と、受取ったキーワードをインデックスファイル23から探し出し、そのキーワードを含むテキストの格納番号をもとにテキスト格納ファイル22から該当テキストを取得する該当テキスト取得手段223とを有して構成されている。

【0003】 この従来の検索方式では、登録時においては、テキスト・キーワード入力手段211により、入力装置21から格納すべきテキストおよび検索時にキーとして使用されるキーワードを入力する。次に、テキスト格納番号取得手段212では、受取ったテキストをテキスト格納ファイル22に格納し、そのテキストをファイル中で一意に決定する格納番号を取得する。次に、インデックスファイル更新手段213では受取ったキーワードと同一のキーワードをもつインデックスレコードをインデックスファイル23から探し出し、そのレコードにテキスト格納番号取得手段212で得られた格納番号を追加し、インデックスファイル23を更新する。検索時においては、キーワード入力手段221により、検索のためのキーとなるキーワードを入力装置21から受取る。次に、テキスト格納番号取得手段222では、受取ったキーワードと同一のキーワードをもつレコードをインデックスファイル23から探し出し、そのレコード中に存在するこのキーワードをキーとするテキストの格納番号を取得する。次に、該当テキスト取得手段223では、テキスト格納番号取得手段222で得られた格納番号をもとに、テキスト格納ファイル22から該当するテキストを取得するようにしている。

【0004】

【発明が解決しようとする課題】 上述した従来の検索方式は、テキスト登録時にテキストをテキスト格納ファイル中で一意に決定する格納番号と、検索のためのキーワードとを直接対応させてインデックスファイルに登録するために、検索時には、そのキーワードと完全に一致した文字列を検索用のキーとして与えない限り、目的とするテキストに付けられた格納番号が得られず、その結果、目的のテキストを取出すことができないという欠点を有している。

【0005】

【課題を解決するための手段】 本発明の検索方式は、キーワードを与えることにより文字列で構成されるテキストを検索する検索方式において、入力装置から文字列で構成されるテキストを受取るテキスト入力手段と、受取ったテキストをテキスト格納ファイルに格納し、そのテキストを一意に決定する番号を取得するテキスト格納番号取得手段と、入力装置から前記テキストを構成するキーワードとその属性とを受取るキーワード・属性入力手段と、このキーワード・属性入力手段が受取ったキーワードと同じ属性を有するレコードをインデックスファイル

ルから探し出し、そのレコードからリンクされているキーワードのレコード群の最後に新たにレコードを追加し、インデックスファイルを更新するインデックスファイル更新手段と、入力装置から検索の対象となるキーワードを受取るキーワード入力手段と、このキーワード入力手段が受取ったキーワードを構成要素とするレコードを前記インデックスファイルから探し出し、そのレコードからリンクされている属性を構成要素とするレコードを取り出し、属性を取得するキーワード検索手段と、このキーワード検索手段で得られた属性をもとに、再度前記インデックスファイルを検索し、同一属性をもつレコードからリンクされている全てのキーワードレコードを探し出し、その各キーワードレコードから更にリンクされている該当テキストの格納番号を構成要素とするレコードから、テキストの格納番号の全てを取得する全該当テキスト番号取得手段と、この全該当テキスト番号取得手段が取得した全ての格納番号をビット列により表現し、論理和をとることにより該当テキストの格納番号を得るビットマップ生成手段と、このビットマップ生成手段で得られたテキスト格納番号をもとに前記テキスト格納ファイルを検索してテキストを取得する該当テキスト取得手段とを有することにより構成される。

【0006】

【実施例】次に、本発明について図面を参照して説明する。

【0007】図1は本発明の一実施例の構成図である。図1(a)はテキスト登録時の構成図で、テキストを入力装置11から受取るテキスト入力手段111と、テキストをテキスト格納ファイル12に格納し、そのテキストをファイル内で一意に決定するテキスト格納番号を取得するテキスト格納番号取得手段112と、入力装置11からテキストを検索するためのキーワードと属性とを受取るキーワード・属性入力手段113と、キーワード属性入力手段113で受取った属性と同じ属性をもつ属性レコードをインデックスファイルから探し出し、キーワードとテキストとの格納番号をそのインデックスレコードからリンクされているレコード群の最後に追加し、インデックスファイルを更新するインデックスファイル更新手段114とから構成される。図1(b)はテキスト検索時の構成図で、検索のためのキーワードを入力装置11から受取るキーワード入力手段121と、キーワード入力手段121が受取ったキーワードを含むキーワードレコードをインデックスファイル13から探し出し、該当レコードにリンクされている属性を構成要素としてもつレコードを取出すキーワード検索手段122と、キーワード検索手段122で得られた属性をもとにインデックスファイル13を検索し、同一の属性を構成要素としているレコードからリンクされている各キーワードのレコードを順次取出し、そのレコードから更にリンクされている該当するテキストの格納番号を構成要素

とするレコードを取出す全該当テキスト格納番号取得手段123と、全該当テキスト格納番号取得手段123で取出された該当する全てのテキスト格納番号からビットマップを生成し、論理演算を施すことにより、取出すテキストの格納番号を入手するビットマップ生成手段124と、ビットマップ生成手段124で入手した該当するテキストの格納番号をもとに、テキスト格納ファイル12からテキストを取出す該当テキスト取得手段125とから構成される。

【0008】図3～図9は図1の実施例の動作を説明するための図で、以下図3～図9を参照して図1の動作について説明を進める。まず図3(a)に示すように、テキスト入力手段111は入力装置11からテキストを受取る。テキスト格納番号取得手段112は受取ったテキストをテキスト格納ファイル12に格納し、ファイル内でそのテキストが一意に決まるように格納された順番に付けられている番号を取得する。図3(a)の例では、初めての格納であるため“1”を格納番号として得ている。次に、図3(b)に示すように、キーワード・属性入力手段113は入力装置11からテキストの検索キーワードとなる文字列とその属性とを受取る。図3(b)の例では“桜”という文字をキーワードとし、その属性には“植物”を採用している場合を示している。なお、ここでいう属性とは、キーワードがもつ意味を広い意味で解釈した場合に、それが属するカテゴリに付けられる名称である。つまり、この例では“植物”的一種として“桜”があることを示し、また“桜”は“植物”という属性をもつことを示している。インデックスファイル更新手段114は、先にテキスト格納番号取得手段112で得られた格納番号と、キーワード・属性入力手段113で得られたキーワードおよび属性とからインデックスファイル13を更新する。

【0009】インデックスファイル13の更新には、まず図4(a)に示すように、属性を構成要素とする属性レコードへのポインタを有する検索レコードがインデックスファイル13中に存在していないために、新たに“植物”をデータとしてもつ属性レコードを生成する。その後、属性レコードへのポインタをデータとする検索レコードを生成する。なお、各レコードの“/”はレコードの終端を示す。次に図4(b)に示すように、“桜”をデータとしてもつキーワードレコードを生成し、その後属性レコードにキーワードレコードへのポインタを格納する。キーワードレコードには、自分自身がリンクされている先のレコードへのポインタも格納する。次に図3(b)のインデックスファイル13に示すように、格納番号“1”をデータとしてもつ格納番号レコードを生成し、その後、キーワードレコードに格納番号レコードへのポインタを格納する。これで1件目のテキストの登録が完了する。

【0010】図5は2件目のテキストの登録によるイン

デックスファイル13の更新の結果の例を示している。図5(a)は2件目のテキストである“北海道では桜と梅がほど同時に開花します”を登録した場合のテキスト格納ファイル12の内容を、図5(b)は図2(a)に対応するインデックスファイル13の状態を示している。ここでキーワードは“桜”および“梅”とし、その属性は共に“植物”としている。2件目のテキストには格納番号として“2”が与えられ、テキスト格納ファイル12に格納する。インデックスファイル13では、“桜”をデータとしてもつキーワードレコードからリンクされている格納番号レコードに“2”を追加し、更に“梅”をデータとしてもつキーワードレコードおよびその格納番号レコードを生成し、リンクする。

【0011】図6はテキストの登録を繰返し、8種類のテキストを登録した場合のテキスト格納ファイル12の状態例を示している。なおテキスト中のアンダーラインで示した語彙はキーワードを示している。図7は図6のテキストの登録におけるインデックスファイル13の状態を示している。

【0012】図8および図9は図6および図7において登録されたテキストから任意のテキストを検索する動作を説明するための図である。まず図8に示すように、キーワード入力手段121において入力装置11から検索のためのキーワードとして“天気”を受取ったとする。キーワード検索手段122では、受取った“天気”をインデックスファイル13から探し出し、該当するキーワードレコードとリンクされている属性レコードから“気象”という属性を取得する。次に図9に示すように全該当テキスト格納番号取得手段123において、再びインデックスファイル7からキーワード検索手段122で取得した属性をデータとしてもつ属性レコードを探し出し、そのレコードからリンクされている全てのキーワードレコードを順次辿り、各キーワードレコードからリンクされている格納番号レコードのデータを取得する。図7に示したインデックスファイル13では“気象”という属性をもつキーワードである“天候”、“天気”、“通り雨”、“雨天”が選択され、その各キーワードレコードから格納番号“3”、“8”、“5”、“7”、“8”が得られる。次に、ピットマップ生成手段124では、各キーワードレコード単位に全テキスト数分のピット数の記憶領域確保し、そのキーワードレコードから

リンクされている格納番号レコード中の番号に対応するピットを“1”にするピットマップを生成し、各キーワードレコード単位にとられたピット列については全体で論理和をとて、重複している番号を一つにまとめる。図7のインデックスファイル13の属性レコード“気象”に基づいて作られた図9に示すピットマップでは“8”が重複しているが、論理和を施すことによって“3”、“5”、“7”、“8”的各ピットが“1”になっている。次に、該当テキスト取得手段125では、各ピットの“1”になっている番号のテキストをテキスト格納ファイル12から取出す。

【0013】

【発明の効果】以上説明したように本発明の検索方式は、キーワードに属性を付加することによって、“天気”、“天候”的に同一の意味をもちながら、従来キーワードが異なると検索できなかつた問題を回避することができるという効果がある。

【図面の簡単な説明】

【図1】本発明の一実施例の構成図である。

【図2】従来の代表的な検索方式の構成図である。

【図3】図1の実施例を説明するための図である。

【図4】図1の実施例を説明するための図である。

【図5】図1の実施例を説明するための図である。

【図6】図1の実施例を説明するための図である。

【図7】図1の実施例を説明するための図である。

【図8】図1の実施例を説明するための図である。

【図9】図1の実施例を説明するための図である。

【符号の説明】

11, 21 入力装置

12, 22 テキスト格納ファイル

13, 23 インデックスファイル

111 テキスト入力手段

112, 212, 222 テキスト格納番号取得手段

113 キーワード・属性入力手段

114, 213 インデックスファイル更新手段

121, 221 キーワード入力手段

122 キーワード検索手段

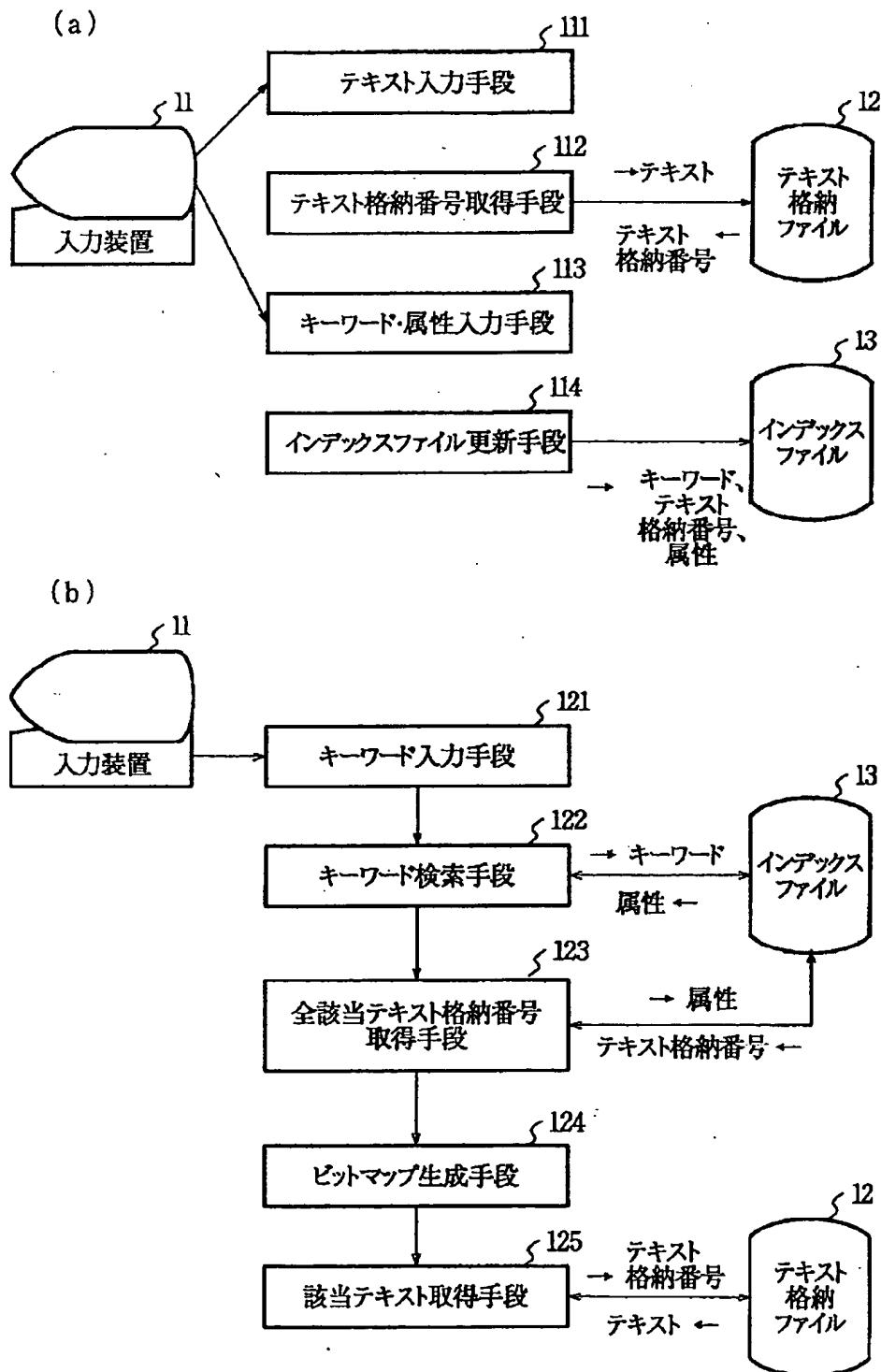
123 全該当テキスト格納番号取得手段

124 ピットマップ生成手段

125, 223 該当テキスト取得手段

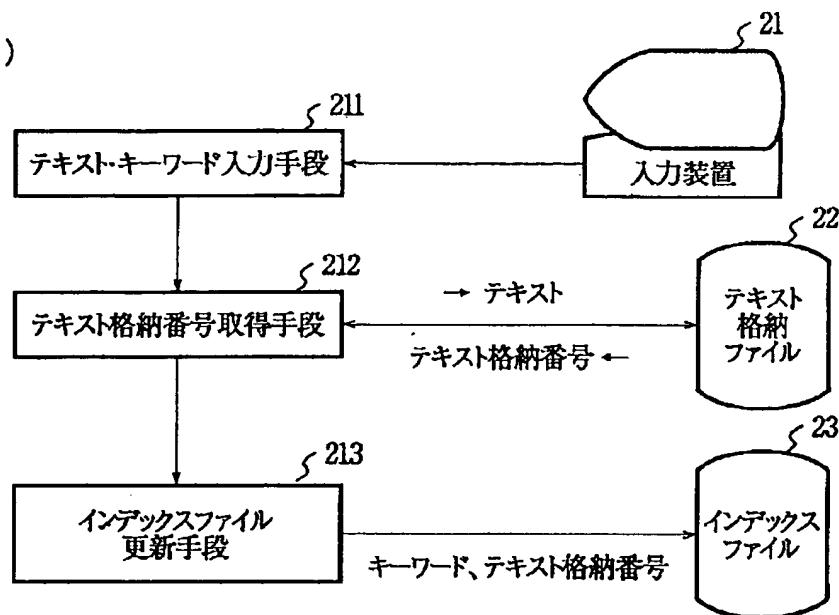
211 テキスト・キーワード入力手段

【図1】

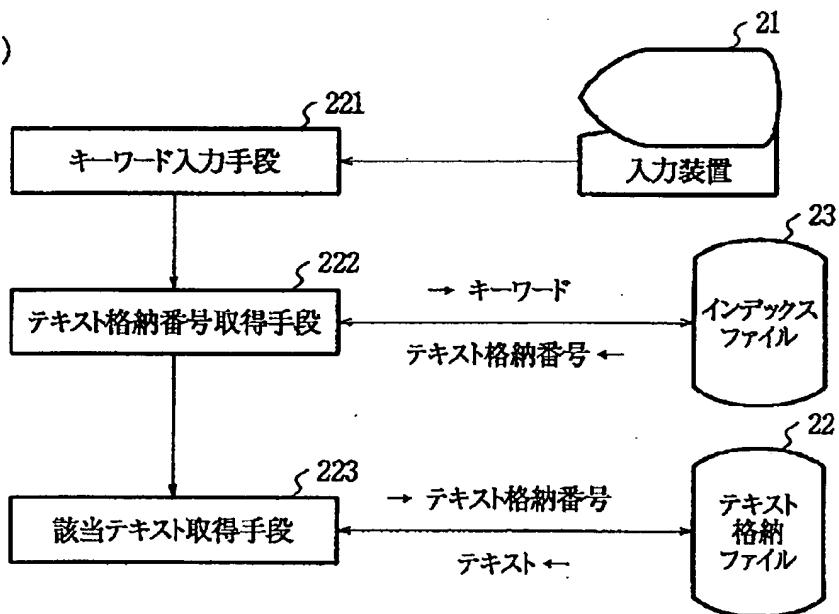


【図2】

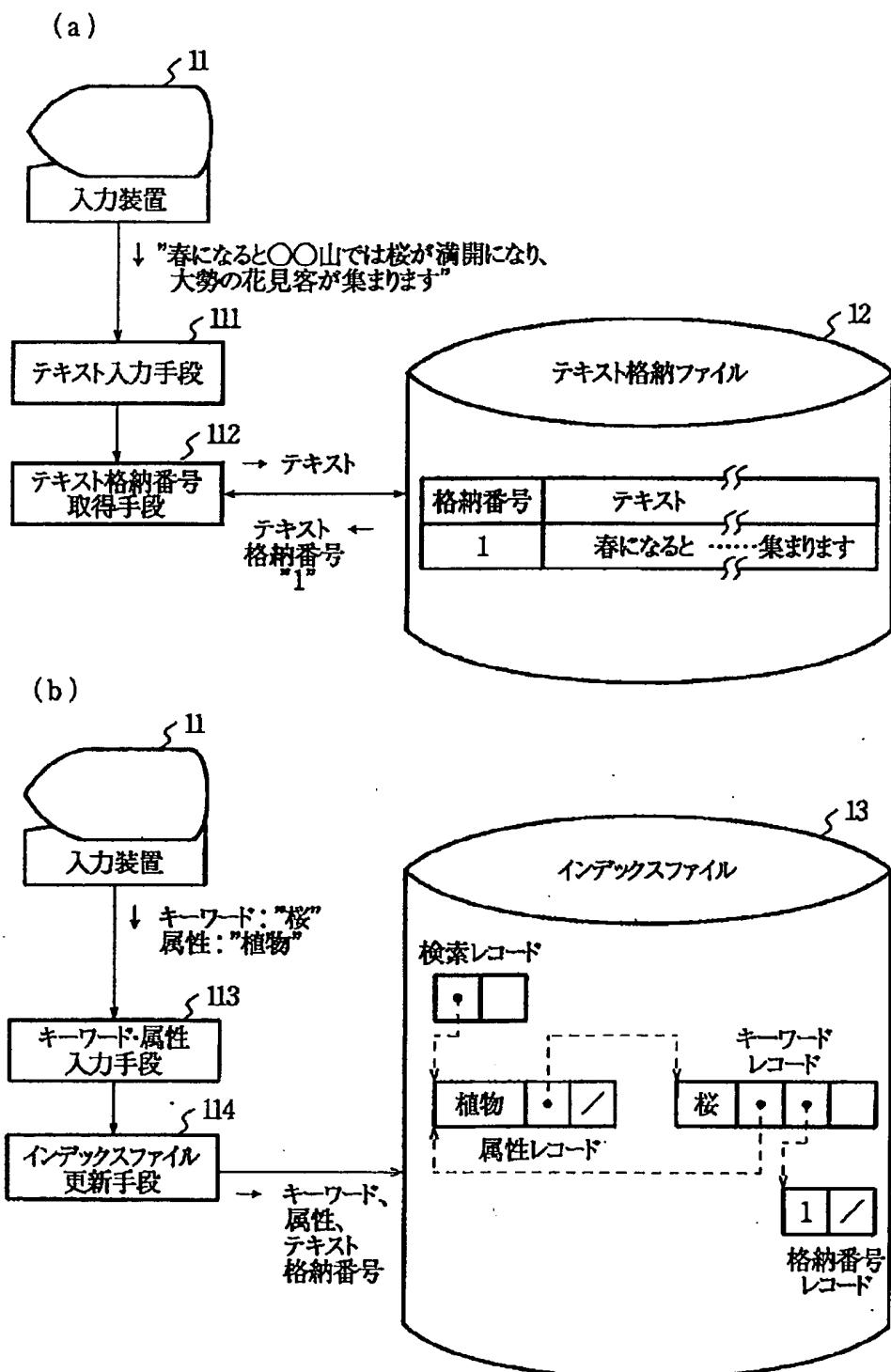
(a)



(b)

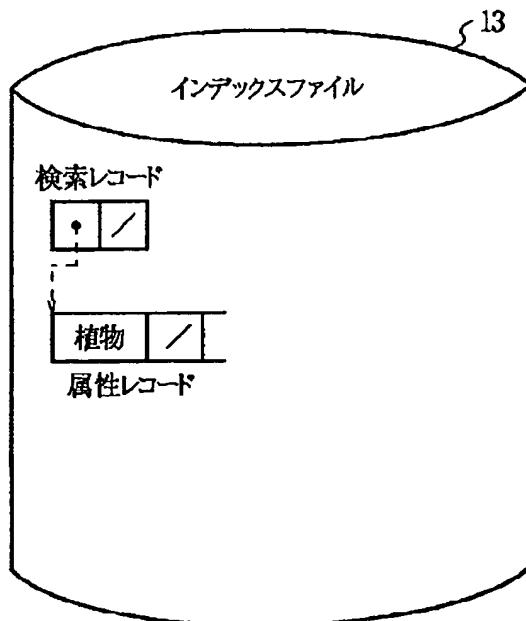


【図3】

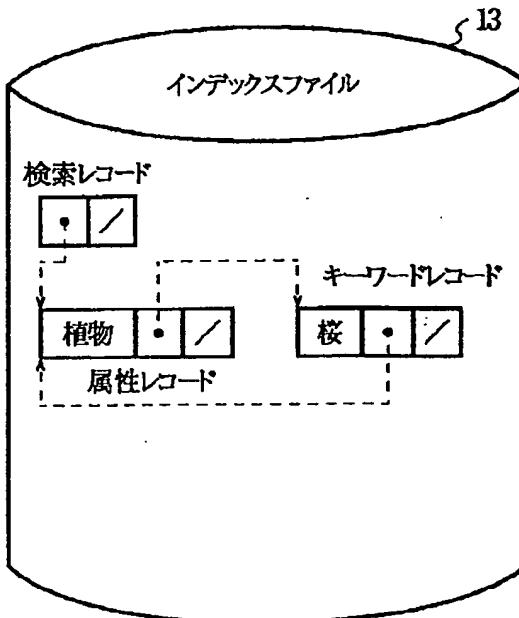


【図4】

(a)

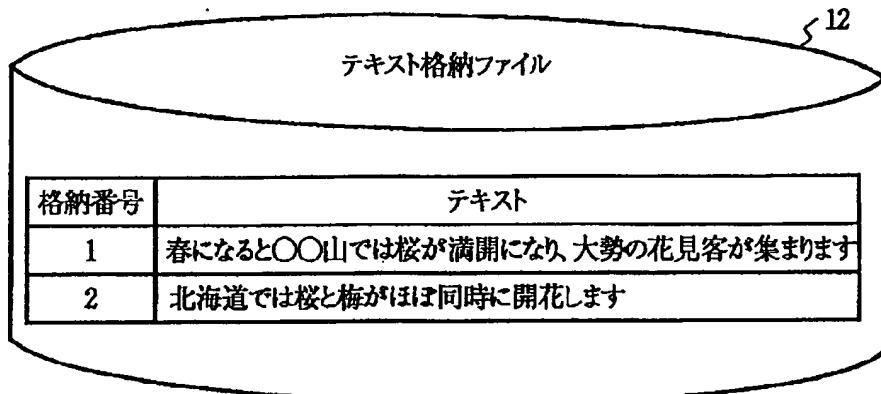


(b)

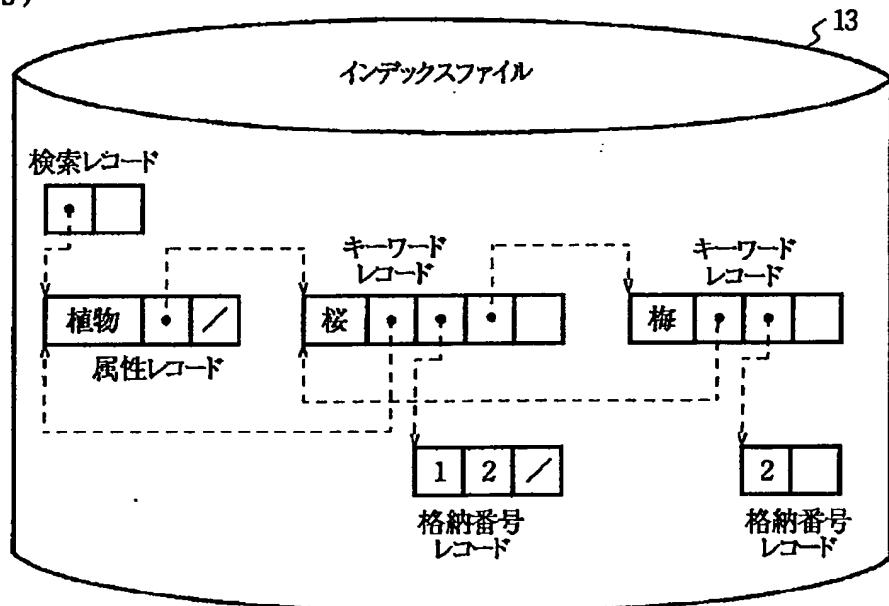


【図5】

(a)



(b)

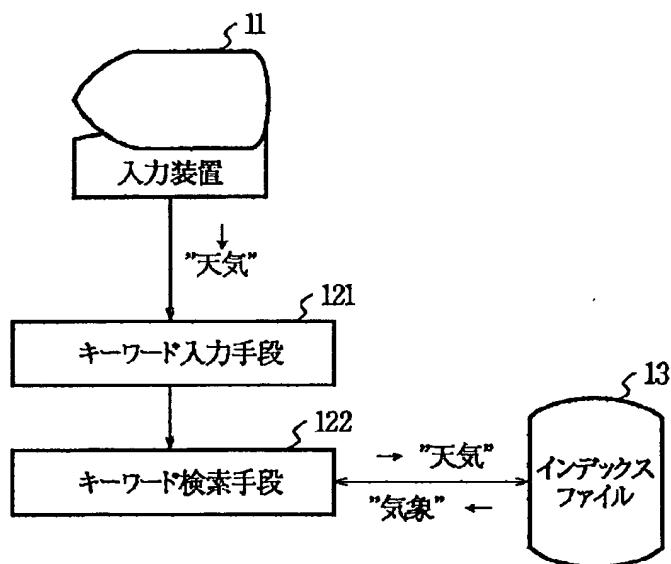


【図6】

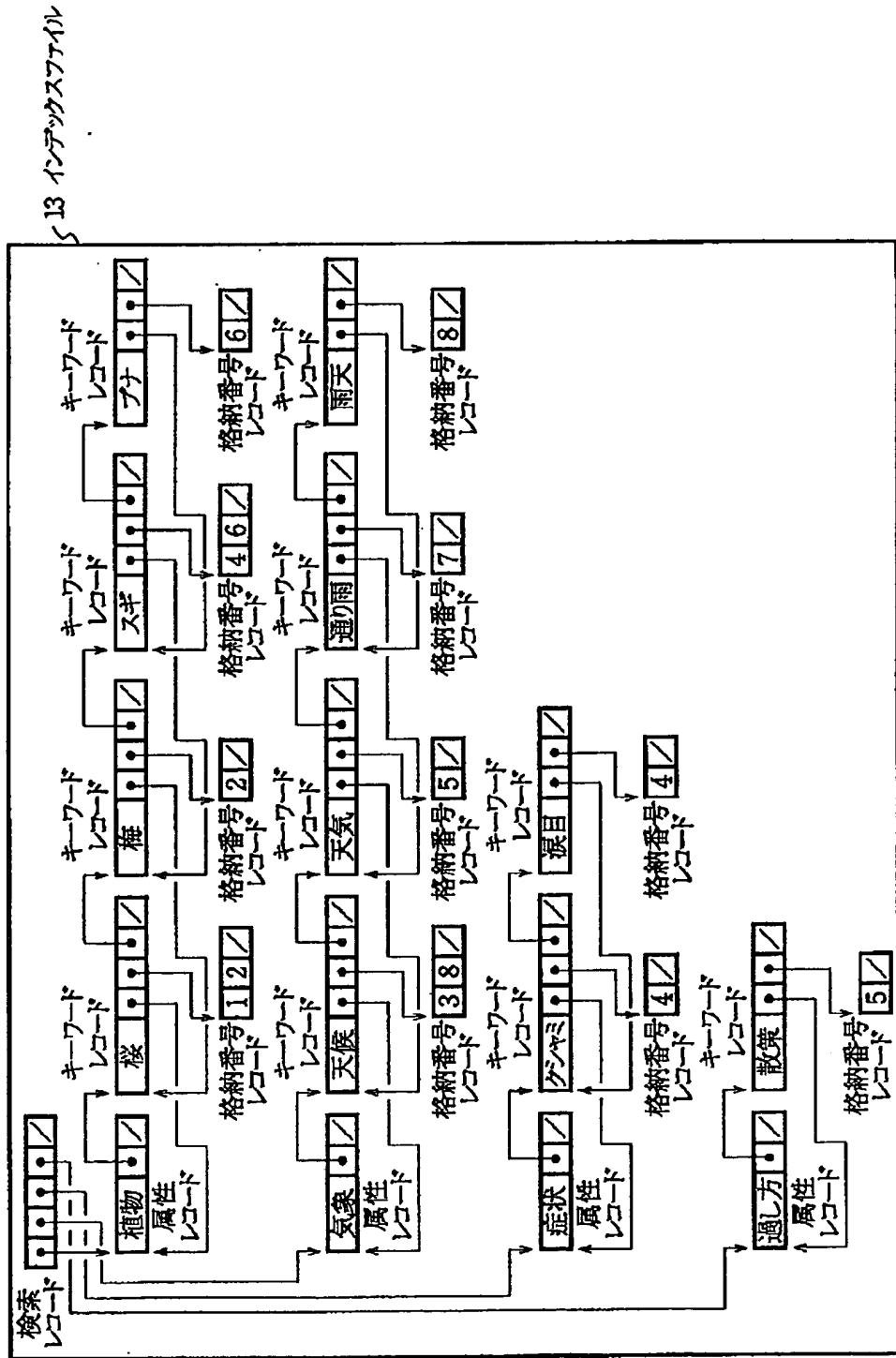
12

テキスト格納ファイル	
格納番号	テキスト
1	春になると〇〇山では <u>桜</u> が満開になり、大勢の花見客が集まります
2	北海道では <u>桜</u> と <u>梅</u> がほぼ同時に開花します
3	XX山は地理的な関係で、頻繁に <u>天候</u> が変わります
4	春先には <u>スギ</u> の花粉により、大勢の人が <u>クシャミ</u> 、 <u>涙目</u> に苦します
5	<u>天気</u> の良い日は散策に出掛けるのもいいでしょう
6	都会に住むと昔見た <u>ブナ</u> や <u>スギ</u> の林が懐かしく思えます
7	夏の暑い中、突然やってくる <u>通り雨</u> は涼しさを残してくれます
8	△△大会は屋内競技のため <u>天候</u> に関係なく、 <u>雨天</u> でも決行します

【図8】



【图7】



【図9】

